Grasas y Aceites
Vol. 50. Fasc. 5 (1999), 369-378

Multivariate models to classify Tuscan virgin olive oils by zone

By **Stefano Alessandri** * [1], **Antonio Cimato** [2], **Giuseppe Modi** [3], **Angela Crescenzi** [4], **Simona Caselli** [5] and **Stefania Tracchi** [3]

[1] Dipartimento di Ortoflorofrutticoltura Via G. Donizetti, 6 - 50144 Firenze, Italy.
E-mail: alessandri@agr.unifi.it
[2] Istituto per la Propagazione delle Specie Legnose - Firenze, Italy.
[3] ARPAT - Dipartimento Provinciale - Firenze, Italy.
[4] Dipartimento Agricoltura e Foreste - Regione Toscana - Firenze, Italy.
[5] Agronomo - Firenze, Italy.

RESUMEN

Modelos multivariantes para clasificar aceites de oliva vírgenes Toscanos por zona.

Para estudiar y clasificar aceites de oliva vírgenes Toscanos, se utilizaron 179 muestras, que fueron obtenidas de frutos recolectados durante la primera mitad de Noviembre, de tres zonas diferentes de la Región. El muestreo fue repetido durante 5 años. Se analizaron ácidos grasos, fitol, alcoholes alifáticos y triterpénicos, dialcoholes triterpénicos, esteroles, escualeno y tocoferoles. Se consideró un subconjunto de variables que fueron seleccionadas en un trabajo anterior como el más efectivo y fiable, desde el punto de vista univariado. Los datos analíticos se transformaron (excepto para el cicloartenol) para compensar las variaciones anuales, restándose la media de la zona Este de los demás valores, dentro de cada año. Se calcularon los modelos de tres clases univariados y además se desecharon variables. Posteriormente, se evaluaron modelos de tres zonas incluyendo fitol (que siempre fue seleccionado) y todas las combinaciones de ácidos palmítico, palmitoleico y oleico, tetracosanol, cicloartenol y escualeno. Se estudiaron modelos incluyendo desde dos a siete variables. El modelo mejor mostró errores de clasificación por zona inferiores al 40%, errores de clasificación por zona dentro del año menores del 45% y errores de clasificación global igual al 30%. Este modelo incluye fitol, ácido palmítico, tetracosanol y cicloartenol.

PALABRAS-CLAVE: Aceite de oliva virgen - Análisis discriminante - Modelo de clasificación - Toscana - Variables canónicas.

SUMMARY

Multivariate models to classify Tuscan virgin olive oils by zone.

In order to study and classify Tuscan virgin olive oils, 179 samples were collected. They were obtained from drupes harvested during the first half of November, from three different zones of the Region. The sampling was repeated for 5 years. Fatty acids, phytol, aliphatic and triterpenic alcohols, triterpenic dialcohols, sterols, squalene and tocopherols were analyzed. A subset of variables was considered. They were selected in a preceding work as the most effective and reliable, from the univariate point of view. The analytical data were transformed (except for the cycloartenol) to compensate annual variations, the mean related to the East zone was subtracted from each value, within each year. Univariate three-class models were calculated and further variables discarded. Then multivariate three-zone models were evaluated, including phytol (that was always selected) and all the combinations of palmitic, palmitoleic and oleic acid, tetracosanol, cycloartenol and squalene. Models including from two to seven variables were studied. The best model shows by-zone classification errors less than 40%, by-zone within-year classification errors that are less than 45% and a global classification error equal to 30%. This model includes phytol, palmitic acid, tetracosanol and cycloartenol.

KEY-WORDS: Canonical variables - Classification model - Discriminant analysis - Tuscany - Virgin olive oil.

1. INTRODUCTION

Tuscany is crossed by a virtual boundary: the northern limit of the cultivation area of *Olea europaea.*

Moreover, the Region shows a great eco-climatic variability, depending on its geographical position and on its complex orography (Maselli *et al.*, 1996, Maracchi *et al.*, 1994).

These two facts make the local production of virgin olive oil very diversified (Alessandri *et al.*, 1997a). However it is extremely difficult to quantify these variations in absolute and reliable terms, because the related influences of the yearly variations of the climatic parameters appear very complex (Alessandri *et al.*, 1997a, Alessandri 1993, Alessandri *et al.*, 1997b). The annual variation of Tuscan olive oils quality can be greater than the variation due to the zone of origin (Alessandri *et al.*, 1997a, Alessandri *et al.*, 1997b). The same can be said for variations due to the harvesting time (Alessandri *et al.*, 1995). Moreover, the interactions among all these factors (Alessandri *et al.*, 1997a) give a difficult resolution to the problem.

Things become more difficult if olive oil characterization and classification is not only a scientific need but also an operational one, expecially if scientific findings have to be the basis for local, regional, national or european regulations.

This work is part of a group of researches to study, and eventually to try and resolve these

problems. Multivariate classification models are discussed, with the aim of classifying Tuscan virgin olive oils by zone. The models are compensated for annual variations. The reliability and the discriminatory power of the variables considered here, was recently analyzed from the univariate point of view (Alessandri *et al.*, 1997a). A multivariate model can show substantial increment of resolution, when compared to the performances of its univariate components.

2. EXPERIMENTAL

2.1. Sampling and experimental design

From the harvest season of 1989-90 to 1993-94 (labeled from «89» to «93» in figures and tables), 179 samples of virgin olive oil were collected. They were obtained exclusively from olives harvested in Tuscany during the first half of November.

In this paper, three zones of Tuscany are considered. The northern zone (labeled N in figures and tables), corresponds to Pistoia, Lucca, and Massa-Carrara. The Western zone (labeled W), covers the Tyrrhenian coast of Tuscany, without its extreme northern part, and extends over Livorno, and a portion of Grosseto. The inner part of Grosseto is included in the Eastern zone (labeled E) with Florence, Arezzo and Siena.

Data related to the N zone and to the 1991-92 harvest season are missing, due to lack of samples.

The descriptive statistics of the samples are reported elsewhere (Alessandri *et al.*, 1997a).

2.2. Chemical Analyses

For each sample fatty acids, Phytol, aliphatic and triterpenic alcohols, triterpenic dialcohols, sterols, Squalene and tocopherols were determined.

Fatty acids were determined using the method stated by the Technical Commission of the Italian Agriculture Ministery (Commissione Tecnica 1976).

Phytol, aliphatic and triterpenic alcohols, triterpenic dialcohols, sterols and Squalene were determined as reported by Modi *et al.* (1991).

Tocopherols were determined by HPLC (RP-18 column, mm 250 length, mm 4,6 diameter; eluting phase of Acetonitrile, Methanol, Acetic Acid 2% water solution; detector UV 294 nm.; I.S. Tocopheril Acetate).

2.3. Statistical Analysis

All the calculations were made by means of the SAS package (Statistical Analysis System).

2.3.1. *Starting Variables*

Among the 34 variables derived from the chemical analyses (Alessandri *et al.* 1997a), only 10 were included in the classification models described in this paper: Palmitic, Palmitoleic and Oleic acid, Phytol Tetracosanol, Hexacosanol, Cycloartenol, Beta-Sitosterol, Delta-5-avenasterol, Squalene.

This first selection follows the assessment of their effectiveness and reliability, that was carried out by means of univariate (Alessandri *et al.*, 1997a) and multivariate (Alessandri *et al.*, 1997b) models. These models were calculated by zone-couples, within each year, on non-transformed data (see below). Data related to the harvest of the 1991-92 growing season were excluded from the calculation of the models, but included in their testing.

2.3.2. *Compensation of yearly variations*

To compensate yearly variation the following steps were taken.

- Data were grouped by year.
- The East zone of Tuscany is the best known in terms of oils (and wines) of high quality and ancient tradition and it is the most studied (Alessandri *et al.*, 1997a). Therefore the East-zone was considered the reference zone. Within each year, all the data were transformed to make the mean of each variable related to the East-zone equal to zero.
- The Cycloartenol values were not transformed, because the «Year» is not a significant source of variation for this variable in Tuscany (Alessandri *et al.*, 1997a).
- Each transformed variable was included in an ANOVA model (Analysis of variance) to confirm that, after the transformation, the «Year» was a not significant source of variation.
- If a variable did not match the condition above, it was discarded.
- The data related to the remaining variables were no more considered as grouped by year and were included in the following calculations as a whole. This approach made a substantial increase of the degrees of freedom of the classification models possible.

2.3.3 *Calculation of the classification models*

All the classification models were based on Linear Discriminant Analysis (Lachenbruch 1975, Hand 1981). They were cross-validated by the leaving-one-out method (Lachenbruch and Mickey 1968), calculated on four-year data (see 3.3.1 and 3.3.2) and then tested on one-year data.

The first calculations were made considering one couple of levels of the Zone variable (W vs. E, W vs. N and E vs. N) at a time. Univariate models were calculated to verify their consistency with the models calculated by year, on non-transformed data (Alessandri *et al.*, 1997a).

Then univariate and multivariate models were calculated, considering the three zones (E, N and W) together. In these models four years of data were included, but their classification errors were analyzed not only as a whole but also within each year (Table I). The variable selection was carried out, by means of the evaluation of all these errors. 50% and 45% of misclassified observations were adopted as selecting thresholds, to evaluate the following:

- The total discriminatory power (total of misclassified observations)
- The capability to recognize the three zones at the same level (observations misclassified by-zone, considering all the years together, except for 1991)
- The yearly consistency (within-year misclassified observations, i.e. total of misclassified observations from the three zones of each single year).

3. RESULTS AND DISCUSSION

3.1. ANOVA

After the transformation described above, the «Year» is not a significant source of variation for any of the 10 variables that were investigated at first (ANOVA calculations not reported). The «Zone», on the contrary, is always highly significant.

3.2. Classification models

The univariate models, calculated by zone-couples (Table I), confirm the results derived from models calculated by year on original values (Alessandri *et al.*, 1997a). Oleic Acid and Phytol score low classification errors for three zone-couples. Palmitic, Palmitoleic acid, and Tetracosanol effectively classify the Western observations vs the others. Hexacosanol and Squalene effectively classify the Eastern oils vs the Western and the Northern.

Table I
Classification errors of the univariate classification models

Analysis Variable	Classification errors (percent)														
	Four year two-class cross-validated models						Four year three-class cross-validated models				Four year three-class mod.: TEST on one-year sub-set of data				
	E	W	E	N	N	W	E	N	W	89	90	(91)	92	93
Palmitic Acid	18	22	40	36	27	30	40	64	30	35	51	(47)	35	55
Total	20		38		29			45						
Palmitoleic Acid	19	26	33	45	27	35	33	73	35	57	39	(48)	45	40
Total	23		39		31			47						
Oleic Acid	18	26	36	32	27	30	36	59	30	56	52	(55)	33	36
Total	22		34		29			42						
Tetracosanol	19	30	39	43	24	30	39	67	30	41	53	(47)	46	44
Total	25		41		27			45						
Hexacosanol	31	39	37	29	52	48	37	81	48	54	58	(58)	65	51
Total	35		33		50			55						
Phytol	11	30	19	33	29	43	**19**	**62**	**43**	**56**	**46**	**(44)**	**15**	**38**
Total	20		26		36			**41**						
Cycloartenol	39	30	26	29	38	39	39	38	70	67	45	(73)	38	62
Total	35		27		39			49						
Beta-Sitosterol	38	43	31	33	43	43	38	43	87	49	55	(64)	55	65
Total	41		32		43			56						
Delta-5-avenasterol	33	43	31	24	48	52	33	48	96	61	54	(67)	59	65
Total	38		27		50			59						
Squalene	35	30	26	24	43	43	35	43	74	64	52	(65)	48	48
Total	33		25		43			51						
Mean										54	51	(57)	44	50

Percent of misclassified observations related to the zones or the years listed below. The data from the 1991-92 harvesting season (Year=91) are not included in all the calculations.
In two-class classification models, the errors greater than 35% are underlined.
In three-class classification models, the errors greater than 45% are underlined and the errors greater than 50% are double-underlined.
All the variables except for Cycloartenol have been transformed (see text).

The three-zone univariate models (Table I) show total classification errors between 41% and 59%. Oleic Acid and Phytol score the lowest errors (42% and 41%), while Delta-5-Avenasterol (59%) e Beta-Sitosterol (56%), the highest.

If we consider the within-year classification errors (excluding 1991), it can be noted that Palmitoleic acid, Tetracosanol and Phytol show only one value greater than 50%. Hexacosanol, Beta-Sitosterol and Delta-5-Avenasterol, on the other side, show three or four values greater than that threshold.

For these reasons Phytol was selected to be included in all the multivariate models, and Beta-Sitosterol, Delta-5-Avenasterol and Hexacosanol were discarded. Therefore the subsequent multivariate models were calculated including all the 63 simple combinations of Phytol and six variables: Palmitic, Palmitoleic and Oleic acid, Tetracosanol, Cycloartenol and Squalene (Tables IIa and IIb).

Table IIa
Classification errors of the multivariate classification models

First Analysis Variable	Second Analysis Variable	Third Analysis Variable	Fourth Analysis Variable	Fifth Analysis Variable	Sixth Analysis Variable	Classification errors (percent)								
						Four-year three-class cross-valid. models				Four-year three-class mod.: TEST on one-year sub-set of data				
						E	N	W	Total	89	90	(91)	92	93
Phytol						19	62	43	41	56	46	(44)	15	38
Phytol	Palmitic Acid					26	62	39	42	53	34	(33)	34	50
Phytol	Palmitol. Acid					26	62	43	44	56	44	(38)	24	49
Phytol	Oleic Acid					24	52	35	37	65	38	(44)	16	28
Phytol	Tetracos.					18	71	48	46	63	46	(38)	16	50
Phytol	**Cycloart.**					**19**	**43**	**48**	**36**	**53**	**39**	**(41)**	**27**	**30**
Phytol	**Squalene**					**23**	**38**	**52**	**38**	**36**	**37**	**(46)**	**32**	**40**
Phytol	Palmitic Acid	Palmitol. Acid				29	57	35	41	53	34	(33)	25	51
Phytol	Palmitic Acid	Oleic Acid				26	48	35	36	70	38	(44)	15	32
Phytol	Palmitic Acid	Tetracos.				23	62	39	41	55	38	(32)	23	50
Phytol	**Palmitic Acid**	**Cycloart.**				**22**	**38**	**39**	**33**	**44**	**26**	**(35)**	**23**	**38**
Phytol	Palmitic Acid	Squalene				29	43	48	40	40	39	(33)	29	43
Phytol	Palmitol. Acid	Oleic Acid				27	57	35	40	56	28	(38)	26	39
Phytol	Palmitol. Acid	Tetracos.				24	71	30	42	55	32	(38)	23	54
Phytol	**Palmitol. Acid**	**Cycloart.**				23	38	39	33	53	31	(46)	12	44
Phytol	Palmitol. Acid	Squalene				26	43	39	36	41	22	(30)	26	47
Phytol	Oleic Acid	Tetracos.				23	57	48	43	63	37	(44)	24	36
Phytol	Oleic Acid	Cycloart.				19	33	35	29	61	36	(46)	11	23
Phytol	Oleic Acid	Squalene				29	38	35	34	51	29	(38)	26	30
Phytol	Tetracos.	Cycloart.				17	48	43	36	53	45	(44)	03	40
Phytol	Tetracos.	Squalene				26	43	43	38	33	37	(38)	26	45
Phytol	Cycloart.	Squalene				23	48	49	40	49	37	(43)	30	30
Phytol	Palmitic Acid	Palmitol. Acid	Oleic Acid			28	57	35	40	55	32	(38)	15	43
Phytol	Palmitic Acid	Palmitol. Acid	Tetracos.			25	67	30	41	55	28	(44)	23	50
Phytol	Palmitic Acid	Palmitol. Acid	Cycloart.			25	43	30	33	44	31	(41)	15	47
Phytol	Palmitic Acid	Palmitol. Acid	Squalene			28	48	35	37	33	17	(24)	29	50
Phytol	Palmitic Acid	Oleic Acid	Tetracos.			25	62	48	45	68	35	(44)	24	40
Phytol	Palmitic Acid	Oleic Acid	Cycloart.			21	33	35	30	61	36	(46)	11	23
Phytol	Palmitic Acid	Oleic Acid	Squalene			27	38	39	35	45	29	(36)	26	30
Phytol	**Palmitic Acid**	**Tetracos.**	**Cycloart.**			**18**	**33**	**39**	**30**	**44**	**31**	**(38)**	**10**	**38**
Phytol	Palmitic Acid	Tetracos.	Squalene			25	43	39	36	28	32	(38)	27	51
Phytol	Palmitic Acid	Cycloart.	Squalene			28	38	43	37	44	33	(38)	24	38

Percent of misclassified observations related to the zones or the years listed below. The data from the 1991-92 harvesting season (Year = 91) are not included in all the calculations.
Errors greater than 45% are underlined. Errors greater than 50% are double-underlined.
For the sake of comparison also the errors related to the Phytol univariate model are reported.

Table IIb
Classification errors of the multivariate classification models

First Analysis Variable	Second Analysis Variable	Third Analysis Variable	Fourth Analysis Variable	Fifth Analysis Variable	Sixth Analysis Variable	Classification errors (percent)								
						Four-year three-class cross-valid. models				Four-year three-class mod.: TEST on one-year sub-set of data				
						E	N	W	Total	89	90	(91)	92	93
Phytol	Palmitol. Acid	Oleic Acid	Tetracos.			23	57	39	40	55	21	(38)	24	43
Phytol	Palmitol. Acid	Oleic Acid	Cycloart.			20	33	35	29	53	25	(16)	12	27
Phytol	Palmitol. Acid	Oleic Acid	Squalene			27	43	35	35	56	17	(36)	28	32
Phytol	Palmitol. Acid	Tetracos.	Cycloart.			20	43	35	33	53	24	(38)	10	49
Phytol	Palmitol. Acid	Tetracos.	Squalene			31	52	35	39	33	18	(38)	28	45
Phytol	Palmitol. Acid	Cycloart.	Squalene			23	43	43	37	48	23	(37)	20	44
Phytol	Oleic Acid	Tetracos.	Cycloart.			21	33	48	34	61	43	(38)	11	31
Phytol	Oleic Acid	Tetracos.	Squalene			29	33	52	38	50	29	(38)	26	38
Phytol	Oleic Acid	Cycloart.	Squalene			24	33	35	31	56	29	(49)	12	23
Phytol	Tetracos.	Cycloart.	Squalene			19	38	48	35	48	37	(48)	21	34
Phytol	Palmitic Acid	Palmitol. Acid	Oleic Acid	Tetracos.		26	62	39	42	60	25	(38)	24	47
Phytol	Palmitic Acid	Palmitol. Acid	Oleic Acid	Cycloart.		20	33	35	29	53	25	(52)	12	27
Phytol	Palmitic Acid	Palmitol. Acid	Oleic Acid	Squalene		26	48	35	36	45	17	(36)	27	32
Phytol	Palmitic Acid	Palmitol. Acid	Tetracos.	Cycloart.		20	33	35	29	53	25	(52)	12	27
Phytol	Palmitic Acid	Palmitol. Acid	Tetracos.	Squalene		28	38	35	34	28	23	(38)	28	50
Phytol	**Palmitic Acid**	**Palmitol. Acid**	**Cycloart.**	**Squalene**		**24**	**43**	**35**	**34**	**40**	**22**	**(38)**	**22**	**43**
Phytol	Palmitic Acid	Oleic Acid	Tetracos.	Cycloart.		21	33	48	34	61	37	(38)	11	31
Phytol	Palmitic Acid	Oleic Acid	Tetracos.	Squalene		28	33	48	37	36	29	(38)	27	38
Phytol	Palmitic Acid	Oleic Acid	Cycloart.	Squalene		24	38	39	34	61	29	(49)	13	23
Phytol	**Palmitic Acid**	**Tetracos.**	**Cycloart.**	**Squalene**		**20**	**43**	**39**	**34**	**40**	**32**	**(35)**	**21**	**45**
Phytol	Palmitol. Acid	Oleic Acid	Tetracos.	Cycloart.		22	33	40	32	53	20	(44)	12	31
Phytol	Palmitol. Acid	Oleic Acid	Tetracos.	Squalene		28	52	43	41	41	17	(38)	27	36
Phytol	Palmitol. Acid	Oleic Acid	Cycloart.	Squalene		23	33	35	30	48	17	(49)	20	36
Phytol	Palmitol. Acid	Tetracos.	Cycloart.	Squalene		22	48	39	36	48	20	(41)	21	44
Phytol	Oleic Acid	Tetracos.	Cycloart.	Squalene		23	38	52	38	56	34	(42)	22	31
Phytol	Palmitic Acid	Palmitol. Acid	Oleic Acid	Tetracos.	Cycloart.	22	38	43	35	53	21	(44)	11	31
Phytol	**Palmitic Acid**	**Palmitol. Acid**	**Oleic Acid**	**Tetracos.**	**Squalene**	**27**	**38**	**43**	**36**	**36**	**18**	**(38)**	**27**	**43**
Phytol	Palmitic Acid	Palmitol. Acid	Oleic Acid	Cycloart.	Squalene	22	38	35	31	56	17	(54)	20	29
Phytol	Palmitic Acid	Palmitol. Acid	Tetracos.	Cycloart.	Squalene	23	43	39	35	40	22	(41)	21	47
Phytol	Palmitic Acid	Oleic Acid	Tetracos.	Cycloart.	Squalene	23	38	52	38	56	29	(41)	22	31
Phytol	Palmitol. Acid	Oleic Acid	Tetracos.	Cycloart.	Squalene	24	38	43	35	48	28	(41)	21	33
All the variables						**24**	**43**	**43**	**37**	**41**	**28**	**(41)**	**22**	**33**

Percent of misclassified observations related to the zones or the years listed below. The data from the 1991-92 harvesting season (Year = 91) are not included in all the calculations.
Errors greater than 45% are underlined. Errors greater than 50% are double-underlined.
For the sake of comparison also the errors related to the Phytol univariate model are reported.
All the variables except for Cycloartenol, have been transformed (see text).

Bivariate models (Table IIa) show that the inclusion of Cycloartenol or Oleic Acid, or Squalene, decreases the total classification error from 41% (Phytol-only model) to the corresponding 36%, 37% and 38%. The by-zone classification errors of the Phytol+Cycloartenol model are all less than 50%. The same can be said about the within-year errors related to the Phytol+Squalene model.

Among the 15 models of three dimensions (Tab. IIa), four lead to a further decrement of the total misclassified observations. Three of them include Cycloartenol and Palmitic, or Palmitoleic, or Oleic Acid. They respectively show errors of 33%, 33%, 29%. The fourth model includes Oleic Acid and Squalene and scores 34% of total classification error. The by-zone classification errors and the within-year errors related to the Phytol+PalmiticAcid+Cycloartenol model are both less than 45%. This particular feature and its effectiveness, both make this the most interesting three-dimension model. It correctly recognizes all the three zones and shows a remarkable stability through the years. This is not the case of the Phytol+Oleic acid+Cycloartenol model. For this reason it was considered less valuable, though it scores a lower total classification error (29%).

Among the 20 models of four dimensions (Tables IIa and IIIb), only the Phytol+Palmitic Acid+Tetracosanol+ Cycloartenol model shows the features emphasized above (Figures 1-5), and a further reduction of the total classification error from 33% (Phytol+Palmitic Acid+Cycloartenol model) to 30%.

Two of the 15 models of five dimensions show within-year errors that match the 45% threshold. Both include Phytol, Palmitic Acid, Cycloartenol and Squalene. The fifth variable is Palmitoleic Acid or Tetracosanol. Both perform worse than the formerly selected models and misclassify 34% of the observations.

This trend is shared by the only six-variable model related to by-zone and within-year errors matching the 45% threshold. It includes Phytol, Palmitic, Palmitoleic, Oleic Acid, Tetracosanol and Squalene, and misclassifies 36% of observations. The same trend is confirmed by the inclusion of the last variable: the seven-dimension model matches the 45% thresholds as above, but misclassifies a further 1% of observations (total classification error 37%).

The items discussed till now can be summarized looking at the performances of the models selected within each number of variables (Table III).

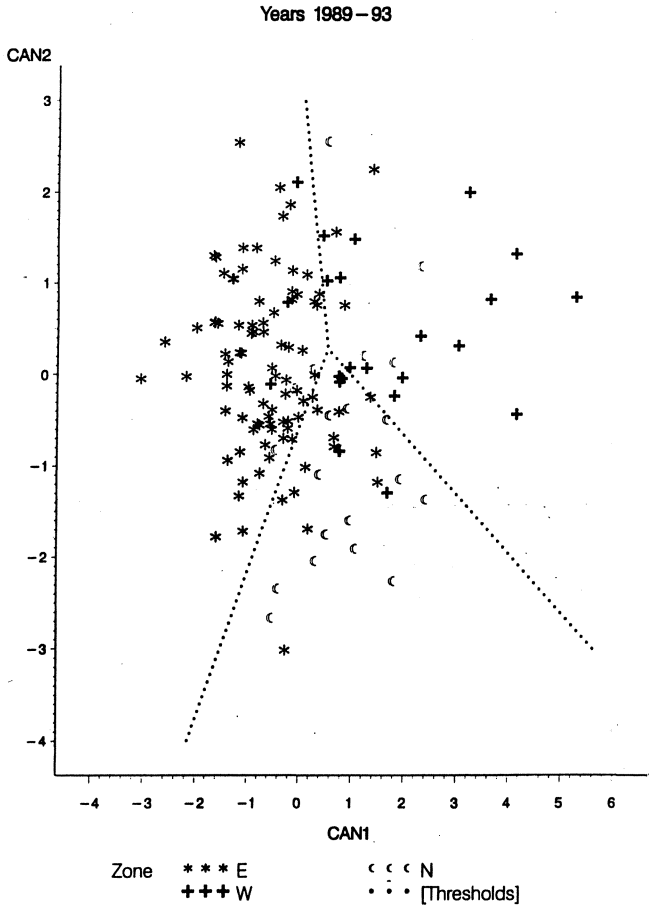

Figure 1
Distribution of the observations in the space of the canonical variates (CAN1 and CAN2). Observations from the Eastern, Northern and Western zones of Tuscany are labeled «E», «N» and «W» respectively. The dotted lines correspond to the between-zone threshold. These are derived from the linear discriminant function including the two canonical variates.
CAN 1 and CAN2 are linear combinations of the transformed (see text) values of Phytol, Palmitic Acid, Tetracosanol, Cycloartenol. The classification model including these variables is the most effective and reliable, among the models considered i this work.
Olive oils extracted from drupes harvested during the seasons of 1989-90, 1990-91, 1992-93 and 1993-94 are represented.

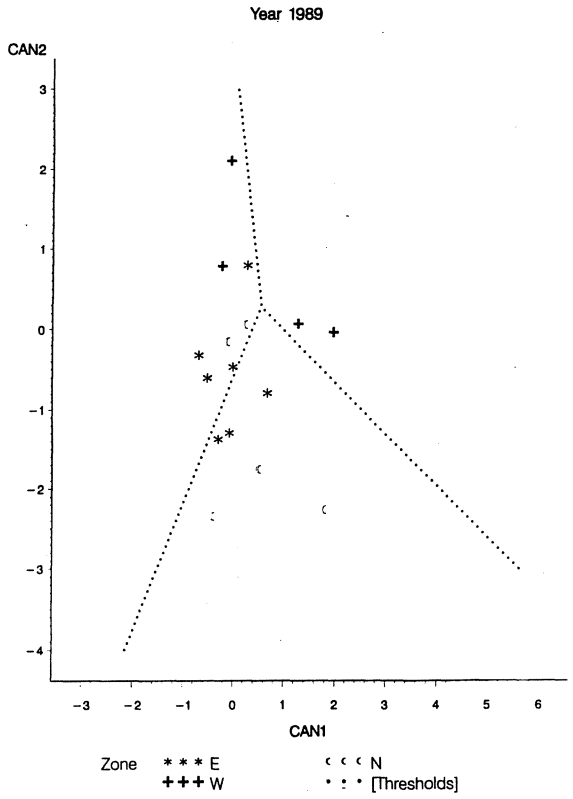

Figure 2
The same classification model as in Figure 1 is shown here, but only Tuscan Olive oils extracted from drupes harvested during 1989-90 are displayed

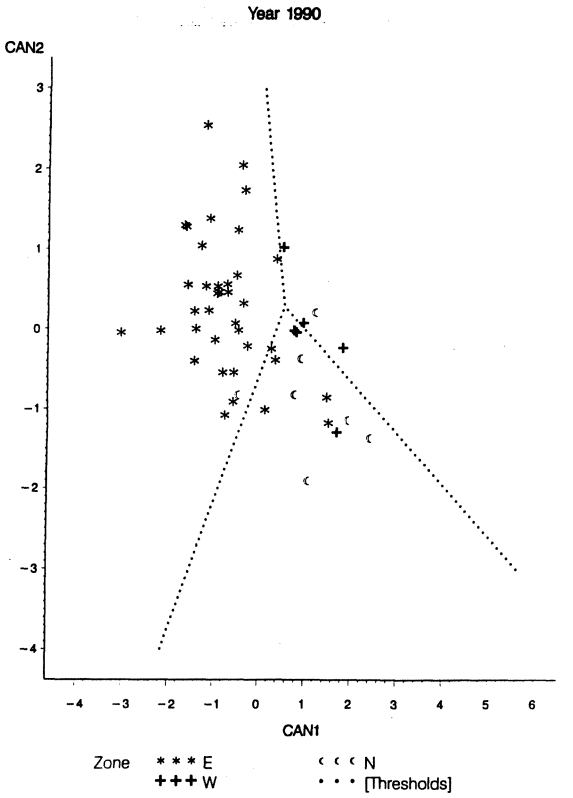

Figure 3
The same classification model as in Figure 1 is shown here, but only Tuscan Olive oils extracted from drupes harvested during 1990-91 are displayed

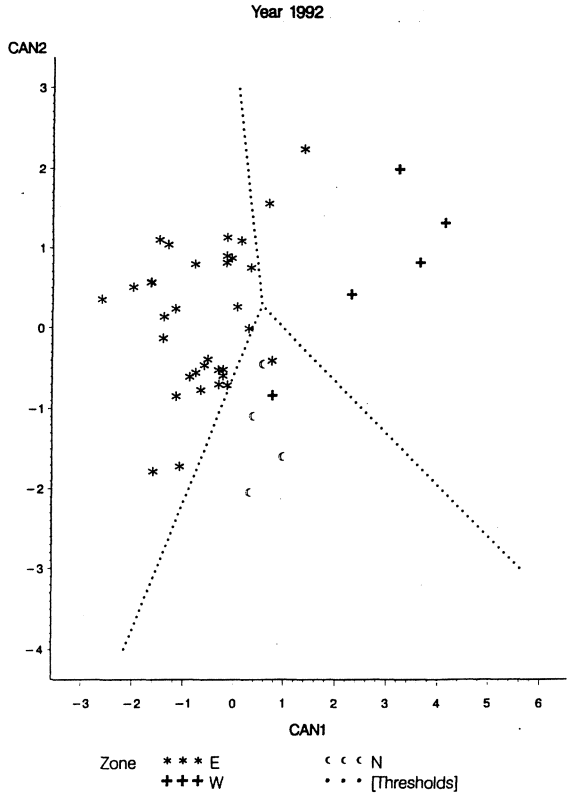

Figure 4
The same classification model as in Figure 1 is shown here, but only Tuscan Olive oils extracted from drupes harvested during 1992-93 are displayed

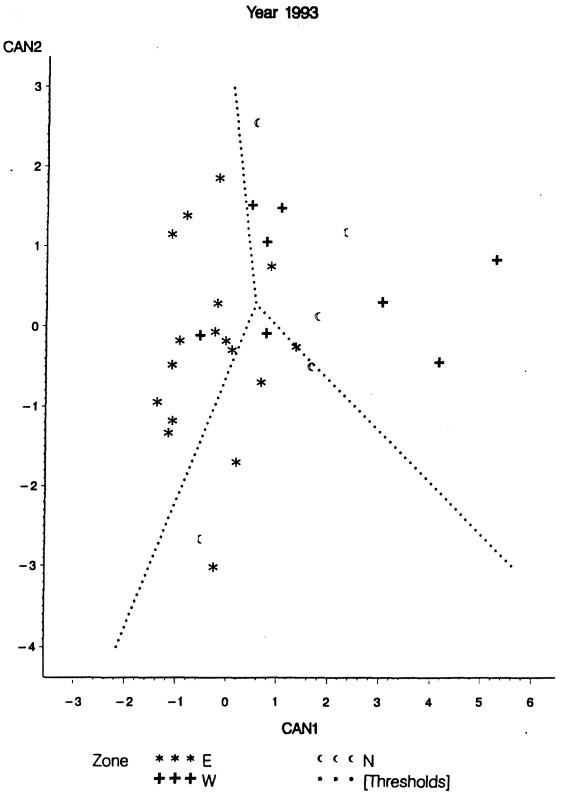

Figure 5
The same classification model as in Figure 1 is shown here, but only Tuscan Olive oils extracted from drupes harvested during 1993-94 are displayed

Table III
Classification errors of the best classification models

First Analysis Variable	Second Analysis Variable	Third Analysis Variable	Fourth Analysis Variable	Fifth Analysis Variable	Sixth Analysis Variable	Classification errors (percent)								
						Four-year three-class cross-valid. models				Four-year three-class mod.: TEST on one-year sub-set of data				
						E	N	W	Total	89	90	(91)	92	93
Phytol						19	62	43	41	56	46	(44)	15	38
Phytol	Cycloart.					19	43	48	36	53	39	(41)	27	30
Phytol	Squalene.					23	38	52	38	36	37	(46)	32	40
Phytol	Palmitic Acid	Cycloart.				22	38	39	33	44	26	(35)	23	38
Phytol	**Palmitic Acid**	**Tetracos.**	**Cycloart.**			**18**	**33**	**39**	**30**	**44**	**31**	**(38)**	**10**	**38**
Phytol	Palmitic Acid	Palmitol Acid	Cycloart.	Squalene.		24	43	35	34	40	22	(38)	22	43
Phytol	Palmitic Acid	Tetracos.	Cycloart.	Squalene.		20	43	35	34	40	32	(35)	21	45
Phytol	Palmitic Acid	Palmitol Acid	Oleic Acid	Tetracos.	Squalene.	27	38	43	36	36	18	(38)	27	43
All the variables						**24**	**43**	**43**	**37**	**41**	**28**	**(41)**	**22**	**33**

Percent of misclassified observations related to the zones or the years listed below. The data from the 1991-92 harvesting season (Year = 91) are not included in all the calculations.
Errors greater than 45% are underlined. Errors greater than 50% are double-underlined.
For the sake of comparison also the errors related to the Phytol univariate model are reported.
All the variables except for Cycloartenol, have been transformed (see text).

Among these nine models only one shows the lowest total classification error of 30%. It is the four-dimension model including: Phytol, Palmitic Acid, Tetracosanol, and Cycloartenol (Table IV, Figure 1-5). It shows a remarkable feature: all its by-zone classification errors are less than 40% (Table III, Figure 1). This characteristic is shared by the selected three-dimension model (Phytol+Palmitic Acid+ Cycloartenol), but this model shows a higher total classification error (33%, Table III). The bi-dimensional representation of the selected four-dimension model (Figures 1-5) was obtained by means of the two related canonical variates. The discriminant linear function and its classifying thresholds were then re-calculated from the canonical variates. The selected four-dimension model was also re-calculated including the non-transformed variables. It is worth underlining that the related total classification error rises to 39% (+9%). This confirms the importance of taking account of yearly variations.

Table IV
Multivariate classification model including Phytol, Palmitic Acid, Tetracosanol, Cycloartenol: Linear Discriminant Function

		Constant = $-.5\,\bar{x}_j$ $COV^{-1}\bar{x}_j$	Coefficient Vector = $COV^{-1}\bar{x}_j$	
Zone		**E**	**N**	**W**
CONSTANT		-12.76260	-7.59981	-11.09198
Phytol	(mg/kg)	0.12775	0.25739	0.29154
Palmitic Acid	(% chrom. area)	-1.15294	-1.02032	0.05379
Tetracosanol	(% chrom. area)	-0.07990	-0.01367	0.12611
Cycloartenol	(% chrom. area)	0.68900	0.51221	0.57362

All the variables except for Cycloartenol have been transformed (see text).

Some results (Table III) reported in this paper can be compared to those regarding univariate models calculated by zone-couples, a year at a time and on original values (3). The two groups of the best performing variables have remarkable overlappings:

- Phytol had been formerly selected to classify Eastern vs Westem and Northern Tuscan olive oils;
- Palmitoleic and Oleic Acid had been formerly selected to classify Western vs Eastern and Northern observations;
- Cycloartenol had been formerly selected to classify Northern vs Eastern and Western oils both in univariate (Alessandri *et al.*, 1997a) and multivariate (Alessandri *et al.*, 1997b) models.

4. CONCLUSIONS

Multivariate effective and reliable models can be calculated, to classify virgin olive oils from Tuscany, on a three-zone basis.

It is important to compensate analytical values for the yearly variations. The method presented here requires either the harvesting year of a blind sample to be known or, for the same year, a reference sampling (the East-zone), to estimate and compensate the variation related to that year.

According to our data, Phytol, Palmitic, Palmitoleic and Oleic acid, Tetracosanol, Cycloartenol and Squalene are related to effectiveness and reliability.

Models including phytol, palmitic acid, cycloartenol with or without tetracosanol misclassify 30% or 33% of the observations. They are related to by-zone classification errors lower than 40% and to within-year classification errors lower than 45%.

ACKNOWLEDGEMENTS

The authors thank the C.R.O.E.V.O.T.T. (Consorzio Regionale Olio Extra Vergine di Oliva Tipico Toscano), the I.V.O.T. (Istituto Vino e Olio di Toscana), the Regione Toscana and all the olive cultivators whose work made this research possible.

The authors are grateful to Marina Spinelli Dondoli for her assistance in editing the text.

BIBLIOGRAPHY

Alessandri, S. (1993).—«Variabilitá annuale e cambiamenti legati al periodo di raccolta delle drupe, nella composizione chimica degli oli toscani».—Proceedings of Presentazione dei resultati delle ricerche inerenti alla caratterizzazione dell'olio extra-vergine di Toscana. Attivitá 1991-92e sviluppi futuri. Appendice. Firenze (Italy) Oct. 19, 1993.

Alessandri, S., Caselli, S., Crescenzi, A., Tracchi, S. (1995).—«Modelli chemiometrici per la classificazione degli oli di oliva in base a cultivar, zona di coltivazione, periodo di raccolta delle drupe».—Bollettino dei Chimici Igienisti parte scientifica, **46**, 405-416.

Alessandri, S,. Cimato, A., Modi, G., Mattei, A., Crescenzi, A., Caselli, S., Tracchi, S. (1997a).—«Univariate models to classify Tuscan virgin olive oils by zone».—La Rivista Italiana delle Sostanze Grasse, **74**, 155-164.

Alessandri, S., Cimato, A., Caselli, S., Baldini, A., Crescenzi, A., Modi, G., Tracchi, S. (1997b).—«The characterization and classification of Tuscan olive oils by zone: yearly variations of the oil composition and reliability of the classification models».—Third International Symposium on Olive Growing, Chania, Greece, 22-26 September, 1997.

Commissione Tecnica Governativa per gli oli minerali, i grassi, i colori, le vernici e i detersivi. (1976).—«Norme italiane per il controllo dei grassi e derivati».—Ministero dell' Industria Roma.

Hand, D. J. (1981).—«Discriminations and classification».—Wiley, New York.

Lachenbruch, P. A. (1975).—«Discriminant analysis».—Hafner Press, New York.

Lachenbruch, P. A., Mickey, M. R. (1968).—«Estimation of error rates in discriminant analysis».—Technometrics, **10**, 1-11.

Maracchi, G., Pittalis, F., Bindi, M., Sillari, B. (1994).—«Olive production e meteorological factors: a preliminary study».—Olivae, **52**.

Maselli, F., Petkov, L., Maracchi, G., Conese, C. (1996).—«Eco-climatic classification of Tuscany through NOAA-AVHRR data».—International Journal of Remote Sensing, **17** (12) 2369-2384.

Modi, G., Fabbrini, A., Simiani, G., Tracchi, S., Mattei, A., Cimato, A. (1991).—«Studio di componenti della frazione insaponificabile di oli toscani ottenuti da varie cultivar».—Bollettino dei Chimici Igienisti, **42**, 787-801.

Recibido: Julio 1998
Aceptado: Enero 1999